1  QUINN EMANUEL URQUHART &
   SULLIVAN, LLP
2  Sean S. Pak (SBN 219032)
   seanpak@quinnemanuel.com
3  50 California Street, 22nd Floor
   San Francisco, CA 94111
4  Telephone: (415) 875-6600
   Facsimile: (415) 875-6700
5
   Andrew H. Schapiro (admitted *pro hac vice*)
6  andrewschapiro@quinnemanuel.com
   191 N. Wacker Drive, Suite 2700
7  Chicago, Illinois 60606
   Telephone: (312) 705-7400
8  Facsimile: (312) 705-4001
9  Alex Spiro (admitted *pro hac vice*)
   alexspiro@quinnemanuel.com
10 295 Fifth Avenue
   New York, NY 10016
11 Telephone: (212) 849-7000
   Facsimile: (212) 849-7100
12
   Rachael L. McCracken (SBN 252660)
13 rachaelmccracken@quinnemanuel.com
   865 South Figueroa Street, 10th Floor
14 Los Angeles, CA 90017
   Telephone: (213) 443-3000
15 Facsimile: (213) 443-3100
16 *Attorneys for Defendant,*
   *NVIDIA Corporation*
17
18                UNITED STATES DISTRICT COURT

19               NORTHERN DISTRICT OF CALIFORNIA

                      OAKLAND DIVISION
20

21 | ABDI NAZEMIAN, BRIAN KEENE, | Master File Case No. 4:24-cv-01454-JST (SK) |
   STEWART O'NAN, ANDRE DUBUS III and
22 SUSAN ORLEAN, individually and on behalf
   of all others similarly situated,          **DEFENDANT'S OPPOSITION TO**
23                                             **MOTION FOR LEAVE TO AMEND**
                      Plaintiffs,              **COMPLAINT**
24
          v.                                   Date:        January 8, 2026
25                                             Time:        2:00 p.m.
   NVIDIA CORPORATION,                         Location:    Courtroom 6
26                                             Judge:       Hon. Jon S. Tigar
                      Defendant.
27

28

# TABLE OF CONTENTS

# **TABLE OF AUTHORITIES**

**Page(s)**

## **Cases**

## Other Authorities

-iii-    Master File Case No. 4:24-cv-01454-JST (SK)

NVIDIA'S OPPOSITION TO MOTION FOR LEAVE TO AMEND COMPLAINT

## I.    INTRODUCTION

Nineteen months after they filed suit and six months after the deadline for amending pleadings, Plaintiffs seek to expand the case in three major ways via their Proposed First Amended Complaint ("PFAC").[1] First, the operative complaint limits the AI models at issue to one specific family, but the PFAC seeks to include any activity related to "LLM training, research, or development" performed by NVIDIA. PFAC ¶ 90. Second, the operative complaint limits the datasets at issue to a subset of The Pile called Books3, but the PFAC now seeks to include "any literary work." *Id*. Third, the operative complaint pled one claim for direct copyright infringement, but the PFAC seeks to add two new claims for vicarious and contributory infringement.

Plaintiffs have failed to show good cause for such a sweeping transformation of the case. Instead, in seeking amendment, Plaintiffs rely solely on information that was available to them before they filed suit, allegations made "on information and belief" that they could have just as easily offered when they filed suit, and irrelevant, conclusory speculation about "shadow libraries" based on unsupported misreadings of a handful of documents produced during discovery.

Plaintiffs' amendment would upend the case schedule and prejudice NVIDIA. NVIDIA has already responded to over 140 written discovery requests, reviewed over three hundred thousand documents and tens of gigabytes of data, and substantially completed its production of non-ESI documents and documents for ESI custodians identified prior to the original substantial completion deadline. Even setting aside the work that has already been done, Plaintiffs offer no plan for how the case could be restarted from the beginning on a massively expanded scope, with only a short time left in fact discovery, much of which is impacted by the holidays. The prejudice associated with this disruption is exactly what the rules are designed to prevent.

Numerous other courts in this district have denied leave to amend or dismissed amended complaints based on allegations like those Plaintiffs seek to make here. *Concord Music Grp., Inc. v. Anthropic PBC*, No. 24-cv-3811 (N.D. Cal. Oct. 18, 2023), ECF 474 at 26:12 (denying leave to

---

[1]  Although Plaintiffs' Exhibit B purports to be a redline of the PFAC over the operative complaint, it includes no redlines. NVIDIA has run a redline, which shows the extensive changes Plaintiffs have made in their PFAC. Declaration of Rachael L. McCracken, ("McCracken Decl.") Ex. 14.

1  amend because the plaintiffs "were not diligent"); *In re Mosaic LLM Litig.*, 2025 WL 2402677, at

2  *2 n.2 (N.D. Cal. Aug. 19, 2025) ("It is the lack of factual, rather than conclusory, allegations as to

3  what materials Databricks actually used to train the DBRX models that dooms Plaintiffs' direct

4  infringement claim against Databricks."); *In re Google Generative AI Copyright Litig.*, 2025 WL

5  2624885, at *7 (N.D. Cal. Sept. 11, 2025) (dismissing Plaintiffs' direct infringement claims as to

6  certain models with prejudice and without leave to amend because Plaintiffs "appear[] to [include]

7  every generative AI model that Google has ever developed" but "do not allege any facts regarding

8  these models at all.").

9          NVIDIA respectfully requests that the Court deny Plaintiffs' motion for leave to amend.[2]

10  **II.    FACT BACKGROUND**

11          **A.  Plaintiffs Filed Their Complaint Nineteen Months Ago**

12          On March 8, 2024, Plaintiffs filed a class action complaint alleging that NVIDIA infringed

13  their copyrights when it trained the NeMo Megatron family of AI large language models ("LLMs")

14  using The Pile, which allegedly contains Plaintiffs' works in a subset called Books3. ECF 1 ¶¶ 19–

15  23; ECF 168. Plaintiffs did not allege their asserted works can be found in any other dataset or that

16  NVIDIA used any other dataset to train the NeMo Megatron LLMs. Although the complaint limited

17  its allegations to the NeMo Megatron family of models, it noted that NVIDIA hosted models "on a

18  website called Hugging Face" (*id.* ¶ 23 n.1), which referred to many other models. After analyzing

19  the disputes at issue in the case, the Court set a case schedule with an April 2025 deadline for

20  amending pleadings (ECF 71) and a March 2026 fact discovery deadline. ECF 188.

21          **B.  The Court Clearly Defined the Scope of Plaintiffs' Case**

22          Despite the clear scope of their complaint, Plaintiffs filed a series of discovery motions

23  seeking to expand the case. Plaintiffs filed a motion seeking discovery on all LLMs "which were

24  trained on or derived from models trained on certain datasets (or "shadow libraries") known to

25  contain copyrighted works," including "The Pile, Books3, Bibliotik, Library Genesis (aka LibGen),

26

27  [2]  The Court should reject Plaintiffs' attempt to dismiss Plaintiff Nazemian under Rule 15; dismissal of a party is governed by Federal Rule of Civil Procedure 41(a) and Plaintiffs failed to respond to

28  NVIDIA's email on conditions for any dismissal.

1  Z-Library (aka B-ok, Sci-Hub, Anna's Archive, or any web scrapes including Common Crawl."

2  ECF 162 at 1. Magistrate Judge Kim rejected Plaintiffs' request and limited discovery to (1) The

3  Pile and (2) LLMs "in the Nemo Megatron family that exist and that trained on the Pile dataset."

4  ECF 168 at 2–3.

5        Plaintiffs sought review of Magistrate Judge Kim's order with the Court, arguing they are

6  entitled to discovery into other datasets because their "asserted works are found not just in Bibliotik

7  and The Pile, but also LibGen, Anna's Archive, and Z-Library" and that these "other pirated libraries

8  are commonly used for LLM development in the AI industry." ECF 174. In affirming Magistrate

9  Judge Kim's order, this Court underscored that discovery must be limited to the "scope of the

10  operative complaint." ECF 183 at 2.

11        Finally, Plaintiffs filed a motion challenging NVIDIA's "definition" of the "NeMo Megatron

12  family," again seeking to expand the scope of the models at issue. ECF 181. Magistrate Judge Kim

13  "decline[d] to expand the universe of models for production by the Defendant" and denied the

14  motion "for the same reasons set forth in the earlier Order." ECF 185 at 2. Plaintiffs did not seek

15  review of that order.

16    **C.  The Close of Fact Discovery Is Rapidly Approaching**

17        During the nineteen months this case has been pending, NVIDIA has reviewed over 240,000

18  custodial documents, over 100,000 non-custodial documents, and many gigabytes of technical

19  materials. McCracken Decl. ¶ 3. NVIDIA has produced 23,467 documents, 69 gigabytes of technical

20  information, and responded to more than 140 written discovery requests. *Id.* ¶ 3. Given Plaintiffs'

21  delays in identifying custodians and pursuing discovery, NVIDIA also needs to review and produce

22  documents for an additional 11 custodians, prepare a privilege log, complete amendment of written

23  discovery responses, and prepare for and defend depositions in the short time left before the March

24  2026 fact discovery cutoff. *Id.* ¶ 4.

25        On October 15, 2025, pursuant to Plaintiffs' request, this Court amended the scheduling

26  order to extend the deadline for substantial completion of document discovery until November 24,

27  2025. ECF 189. But that extension was requested solely to allow Plaintiffs more time to complete

28

1    their own production and complete their custodian identification. ECF 186 at 3. The order did not

2    address Plaintiffs' efforts to expand the case. ECF 189.

3    **III.    LEGAL STANDARD**

4        Once a district court enters a pretrial scheduling order in a case, modifications require "good

5    cause." Fed. R. Civ. P. 16(b)(4). A party moving to amend must "demonstrate that amendment [is]

6    proper under Rule 15." *Johnson v. Mammoth Recreations, Inc.*, 975 F.2d 604, 607-08 (9th Cir.

7    1992). And the Court may "exercise its discretion to deny leave to amend" for a variety of reasons,

8    including "undue delay," "undue prejudice to the opposing party," and "futility of amendment."

9    *Carvalho v. Equifax Info. Servs., LLC*, 629 F.3d 876, 892 (9th Cir. 2010) (quoting *Foman v. Davis*,

10    371 U.S. 178, 182 (1962)).

11    **IV.    ARGUMENT**

12        **A.  <u>Plaintiffs Have Not Shown Good Cause to Expand the Case from One LLM Family</u>**

13            **<u>to All of NVIDIA's LLM Research, Development, and Training</u>**

14        Plaintiffs' proposed amendment seeks to convert a case about a single family of LLMs into

15    an open-ended fishing expedition into all the work NVIDIA has done on LLMs. As Magistrate Judge

16    Kim has previously ruled, the operative complaint limits the scope of the models in the case to "the

17    NeMo Megatron family" of LLMs. ECF 168 at 3:11-13; ECF 185 at 2:3-5 ("The Court declines to

18    expand the universe of models[.]"). Now, the PFAC seeks to expand the class definition to include

19    anything related to any "LLM training, research, or development" by NVIDIA. *Compare* PFAC

20    ¶ 90 (class definition), *with* ECF 1 ¶ 39 (class definition). The PFAC adds new allegations referring

21    generically to NVIDIA's "language models" (PFAC ¶¶ 4-5, 9, 66, 78) and NVIDIA's "LLMs" (*id.*

22    ¶¶ 20, 74-75, 78). It even seeks to expand the case to include "internal research which did not result

23    in a fully trained LLM" (*id.* ¶ 43); "internal models, including checkpoints," that were not published

24    or even named (*id.* ¶ 42); and other undefined aspects of the "LLM development process" (*id.* ¶ 22).

25    Thus, although Plaintiffs' motion insinuates that the PFAC would be limited to LLMs "trained on

26    the Pile" (Mot. at 2), their proposed allegations go much further.

27        Plaintiffs fail to show good cause for such a sweeping reformulation of the case. The good

28    cause standard "primarily considers the diligence of the party seeking the amendment." *Johnson*,

NVIDIA'S OPPOSITION TO MOTION FOR LEAVE TO AMEND COMPLAINT

975 F.2d at 609. Where Plaintiffs are "aware of a possible theory but fail[] to timely raise it or alternatively provide a good reason for not doing so, the court is unable to find that Plaintiff[s] acted with diligence." *Gupta v. Int'l Bus. Machs. Corp.*, 2015 WL 9204348, at *2 (N.D. Cal. Dec. 17, 2015).

Here, the fact that NVIDIA was working on models other than the NeMo Megatron family was public long before Plaintiffs filed their original complaint, as Plaintiffs' own allegations show. For example, both the operative complaint and the PFAC note that NVIDIA posted models publicly on a prominent AI website called "Hugging Face." ECF 1 ¶ 23; PFAC ¶ 29. The operative complaint cites four models from NVIDIA's Hugging Face page, but as of February 28, 2024, NVIDIA's Hugging Face page included 7 collections and 126 models. McCracken Decl. Ex. 1; *see also* McCracken Decl., Ex. 2 (showing the release date and total downloads of 124 NVIDIA models that pre-date Plaintiffs' original complaint); McCracken Decl., Ex. 3 (showing the information about various NVIDIA models that was publicly available before Plaintiffs filed their original complaint).

In addition to their conclusory allegations about NVIDIA's work on LLMs generally, Plaintiffs point to five specific models. But information about those models was also publicly available before the case started.[3] Plaintiffs identify the "Retro" models (InstructRetro-48B, Retro-48B) as models outside the NeMo Megatron family that should be swept into the case. PFAC ¶ 40. Those models were described in a publication in October 2023, long before the complaint's March 2024 filing date and the April 2025 deadline for amending the pleadings. McCracken Decl., Ex. 4 at 15. The publication not only described the "Retro" models but also stated that they were trained using "a pretraining dataset consisting of … book data (e.g., Book3 and Book-Corpus2 from the Pile dataset …)." *Id.* In fact, Plaintiffs' Exhibit I recognizes that the "InstructRetro paper" was "getting a lot of attention" in November 2023 and quotes the publication's training-data disclosure. Pls. Ex. I at 2, 4.

---

[3]  Plaintiffs also identify "███████████" as supporting their proposed amendment. Pls. Ex. D. But that model already falls within the scope of Plaintiffs' operative complaint as defined by Magistrate Judge Kim, because it is an LLM in the NeMo Megatron family that was trained on The Pile.

NVIDIA'S OPPOSITION TO MOTION FOR LEAVE TO AMEND COMPLAINT

1    Similarly, Plaintiffs cite the Nemotron-4 340B and 15B models, speculating that those

2  models must have been trained using Plaintiffs' works because they were trained with a large

3  dataset. PFAC ¶¶ 58-60. The technical paper that Plaintiffs rely on for their speculation was released

4  on February 27, 2024—before Plaintiffs filed suit and more than a year before the deadline for

5  amending the pleadings. McCracken Decl., Ex. 5.

6    Finally, Plaintiffs cite the Megatron GPT2 345M model (PFAC ¶ 40), but that model was

7  disclosed in a September 2019 publication, more than four years before Plaintiffs filed suit.

8  McCracken Decl., Ex. 6; *see also id.*, Ex. 7 (confirming that this model was publicly available before

9  Plaintiffs filed suit). Plaintiffs point to a document produced in discovery stating that this model was

10  trained on a ███████████████████████████████████." McCracken Decl., Ex. 8. Plaintiffs

11  misread this document to suggest that the model was trained on The Pile itself. To the contrary, that

12  document says only that the model was trained on certain resources, some of which are also used in

13  The Pile—as disclosed in the September 2019 publication—including Wikipedia, OpenWebText,

14  RealNews, and CC-Stories.[4] *Id.* Books3, the sole component alleged to contain Plaintiffs' works,

15  was ***not used***. *Id.* Regardless, that document does not justify expanding the case to include a model

16  that was disclosed in 2019, let alone to include all of NVIDIA's LLM research, development, and

17  training activities.

18    Plaintiffs cannot show diligence, because the information underlying the amendments was

19  publicly available before the complaint was filed, nineteen months ago. Plaintiffs' long delay is well

20  beyond the "eight month delay between the time of obtaining a relevant fact and seeking a leave to

21  amend" that the Ninth Circuit has held to be "unreasonable." *AmericsourceBergen Corp. v. Dialysist*

22  *W., Inc.*, 465 F.3d 946, 951 (9th Cir. 2006); *Kamal*, 88 F.4th at 1277–79 ("eight month[]" and "five

23  month[]" delays dilatory); *M.A. Silva Corks USA, LLC v. M.A. Silva Holdings, Inc.*, 2024 WL

24  1180999, at *4–7 (N.D. Cal. Mar. 18, 2024) (no diligence where party knew facts six months before

25  deadline but waited to move until after it had passed).

26

_____

27  [4] Plaintiffs' Exhibit D is an excerpt of a spreadsheet that does not mention the Megatron GPT2 345M, but the full spreadsheet produced by NVIDIA says that the model was trained with a ██████

28  ██████████ that excludes Books3. McCracken Decl., Ex. 8.

1    *Bot M8 LLC v. Sony Corp. of Am.*, 2020 WL 1643692 (N.D. Cal. Apr. 2, 2020) is instructive.

2    There, the court denied leave to amend where the plaintiff sought to add allegations of infringement

3    "entirely based on public information from the web" that were available before filing suit. *Id.* at *3.

4    The court emphasized that because the plaintiff "controlled the timing of this case" and "decided

5    when to sue," the allegations "should have been included in the operative complaint" and "certainly

6    should have been included" when the plaintiff amended its complaint as of right. *Id.*; *Baker v. Santa*

7    *Clara Univ.*, 2018 WL 3491689, at *3 (N.D. Cal. July 19, 2018) (denying leave to amend where

8    "the proposed amendments [we]re based on information that Plaintiff had or should have had before

9    filing suit"); *Kaplan v. Rose*, 49 F.3d 1363, 1370 (9th Cir. 1994) (affirming denial of request for

10   leave where information "w[as] known to [Plaintiffs] from the beginning of the litigation").

11   Similarly here, the fact that NVIDIA was training models other than those in the NeMo Megatron

12   family, including on literary texts, was publicly known before the complaint was filed. If Plaintiffs

13   wanted to pursue a broader case, they needed to do so when they filed their original complaint, not

14   nineteen months into the litigation and six months after the deadline for amending.

15       Plaintiffs' cited cases do not hold otherwise. *Lee v. Amguard Insurance Company* found

16   good cause to amend after a delay of only three months between production and amendment due to

17   "the challenges caused by the pandemic" and interference of "COVID-19 and stay at home

18   restrictions" on plaintiffs' counsel's ability "to meet with and discuss this case with his clients."

19   2021 WL 3161683, at *6 (N.D. Cal. July 27, 2021). Similarly, *Juarez v. Auto Zone Stores, Inc.*

20   found good cause to amend because plaintiff moved for leave one business day after he learned

21   pertinent information that defendant withheld during discovery. 2010 WL 3470759, at *2 (S.D. Cal.

22   Sept. 3, 2010). None of those circumstances applies here. Instead, because Plaintiffs "w[ere] not

23   diligent," the Court should deny Plaintiffs' motion. *Kamal*, 88 F.4th at 1277 (quoting *Johnson*, 975

24   F.2d at 609).

25       In addition to lacking diligence, Plaintiffs' new allegations about "NVIDIA's LLMs" and its

26   internal research activities are too conclusory to state a claim for copyright infringement. Apart from

27   the five specific models addressed above, Plaintiffs allegations refer to "NVIDIA's LLMs" and its

28   "internal research" only in the abstract. *E.g.*, PFAC, ¶¶ 4-5, 9, 20, 42-43, 66, 74-78. Plaintiffs

1    "cannot use one alleged instance of infringement to simply presume that the same party committed

2    more infringement." *In re Mosaic LLM Litig.*, 2025 WL 2402677, at *3. Here, Plaintiffs do not offer

3    any plausible allegations that all of NVIDIA's LLMs or all its research, development, or training

4    activities necessarily used Plaintiffs' works. And "simply speculating on how widespread the

5    copying may be, without describing any acts or works that infringe, fails to state a claim for direct

6    copyright infringement." *Id.* In short, Plaintiffs do nothing more than "presume" NVIDIA's LLMs

7    trained on the cited datasets, which is not sufficient to state a claim. *Andersen*, 700 F. Supp. 3d at

8    868–69.

9          **B.**   **Plaintiffs Have Not Shown Good Cause to Add Any New Datasets, Let Alone**

10              **Expand the Case to Cover All Text-Based Works**

11        Plaintiffs' proposed amendment also seeks to convert a case about a single dataset Plaintiffs

12    alleged contains their works into a fishing expedition into any "literary work." As Magistrate Judge

13    Kim previously ruled, the operative complaint limits the scope of the discovery to "training on The

14    Pile." ECF 168 at 3:3-5. The Court held that the order properly "limited the scope of the databases

15    that the Plaintiffs could examine to only the Books3 dataset that Plaintiffs specifically alleged as the

16    source of the pirated training material." ECF 183 at 3:16-28. Now, the PFAC seeks to expand the

17    class definition to encompass "any literary work." PFAC ¶ 90 (class definition). The PFAC would

18    also add allegations that NVIDIA used four specific datasets and "other shadow libraries." *Id.* ¶¶ 74-

19    76.

20        The Court should also reject this effort to expand the case for several reasons. ***First***, the

21    PFAC fails to allege that any of the new datasets were used to train the NeMo Megatron family of

22    models or any specific models. As discussed in the preceding section, the operative complaint is

23    limited to a single family of models, and Plaintiffs have failed to show good cause for expanding

24    the accused models now. Because the PFAC does not provide any link between the new datasets

25    and the only models at issue in this case, or any specific models at all, the new dataset allegations

26    have no relevance to the case and any amendment would be futile. *Ahlmeyer v. Nev. Sys. of Higher*

27    *Educ.*, 555 F.3d 1051, 1055 (9th Cir. 2009); *Miller v. Calhoun Cty.*, 408 F.3d 803, 817 (6th Cir.

28    2005); *Colley v. Vierra*, 2023 WL 9319113, at *5 (C.D. Cal. Dec. 18, 2023).

1    **Second**, although Plaintiffs' motion emphasizes ███████████████████████ is **not a**

2    **dataset**; it is a ███████████████████████████████████████████████████

3    ██████—including many relating to academic books, academic papers, foreign language materials,

4    and other categories of data that the PFAC does not accuse. Tellingly, Plaintiffs do not tie any

5    allegations about ███████████ to the claims at issue, nor do they make any specific allegation

6    about it. For example, the PFAC lacks any factual allegation that NVIDIA used ████████████

7    ██████  The Pile, or any other dataset alleged to contain Plaintiffs' works. The sole "specific

8    information" cited is a single email that indicates only that an NVIDIA research team was "████████

9    █████████████████████████████" (Mot. at 3-4) but does not state that NVIDIA actually

10   acquired any datasets █████████████████████████████, much less acquired The Pile

11   or any other dataset that Plaintiffs newly allege contained their works ████████████, or used

12   any dataset ████████████████████ to train any LLM, much less a specific LLM. Plaintiffs

13   are essentially making all factual allegations core to the claims on information and belief—

14   something they could have done at any time, including when they originally filed. Plaintiffs'

15   allegations about ███████████ also fail to "to rise above the speculative level." *Bell Atlantic*

16   *Corp. v. Twombly*, 550 U.S. 544, 555 (2007). Instead, the email does nothing to confirm Plaintiffs'

17   suspicion that NVIDIA might have used datasets ████████████████████████████████

18   beyond what Plaintiffs knew from publicly available sources that precede the complaint, including

19   their own court filings in this, *see* ECF 1 ¶ 27 (identifying ███████████ as a "notorious 'shadow

20   library'"), and other cases, *see, e.g.*, *Tremblay v. OpenAI, Inc.*, No. 23-cv-03223 (N.D. Cal. June 28,

21   2023), ECF 1 ¶ 34; *Kadrey v. Meta Platforms, Inc.*, No. 23-cv-03417 (N.D. Cal. July 7, 2023), ECF

22   1 ¶ 27; *O'Nan v. Databricks, Inc.*, No. 24-cv-01451 (N.D. Cal. Mar. 8, 2024), ECF 1 ¶ 28.

23   **Third**, Plaintiffs cite a Hugging Face page about SlimPajama. PFAC ¶ 41 n.8 (citing

24   https://huggingface.co/datasets/cerebras/SlimPajama-627B). Highlighting Plaintiffs' lack of

25   diligence, this page was publicly available when the original complaint was filed. McCracken Decl,

26   Ex. 9 (archive of the SlimPajama dataset page as of January 2024). Moreover, although the page

27   lists models that were supposedly trained using SlimPajama, the list does not include any NVIDIA

28   models, let alone the ones at issue in this case. https://huggingface.co/datasets/cerebras/SlimPajama-

627B. McCracken Decl. ¶ 15. Plaintiffs also point to a document from discovery indicating that

NVIDIA asked to borrow an ███████████████ ) to test a product from a third-party, ██ .

Mot. at 11 (citing Pls. Ex. N). But that email shows that the ████ project was ██████████

██████████████ —a fact Plaintiffs omit from their Motion. Pls. Ex. N ███████████████

████████████████████████████████████████████████████████████████████

██████████████████████████████████████████████ (emphasis in

original).[5] Plaintiffs have thus failed to plead "non-speculative facts about what Defendant infringed

and how." *In re Mosaic LLM Litig.*, 2025 WL 2402677, at *2 (N.D. Cal. Aug. 19, 2025).

*Fourth*, the new allegations as to three of the four datasets (LibGen, Sci-Hub, and Z-library)

are pled entirely on "information and belief," without citing to any supporting discovery. PFAC

¶ 76. If Plaintiffs intended to plead these dataset allegations based purely on "information and

belief," they could have included them in their original complaint, where all three were identified

as "notorious 'shadow libraries'" (ECF 1 ¶ 27), or amended before the April 2025 deadline.

Moreover, "even allegations on information and belief must be supported by a factual foundation

leading to that belief." *In re Mosaic*, 2025 WL 2402677, at *2 n.2 (cleaned up). Here, "the lack of

factual, rather than conclusory, allegations as to what materials [NVIDIA] actually used to train [its]

models" renders any proposed amendment as to these datasets futile. *Id.*

Accordingly, the Court should reject Plaintiffs' efforts to expand the datasets, because their

allegations are untimely, conclusory, and do not tie those datasets to the models at issue.

**C. Plaintiffs Have Not Shown Good Cause to Add New Indirect Infringement Claims**

Finally, Plaintiffs have also failed to show good cause to add two new counts to the

complaint. The operative complaint pleads only one claim, a count for direct copyright infringement.

ECF 1 ¶¶ 32-37. The PFAC seeks to add two new claims, a count for contributory infringement and

a count for vicarious copyright infringement. PFAC ¶¶ 81-88.

---

[5] Plaintiffs' Exhibit N does not relate to using SlimPajama or any dataset to train LLMs—the alleged infringing activity. Rather, it relates to ████████████████████████ , which is unrelated to the issues in either the original or proposed complaints.

The new counts are based on conclusory allegations about NVIDIA's "NeMo Framework," which Plaintiffs mislabel as the "NeMo Megatron Framework." PFAC ¶¶ 67-69. Plaintiffs cite a YouTube clip from March 2023, describing the framework as a wide-ranging set of tools to help developers in creating their own models. *Id.* ¶ 67 n.10. The PFAC alleges that one feature included in the framework was an optional script for downloading The Pile. *Id.* ¶ 68. The PFAC then speculates "on information and belief" that NVIDIA provided the framework to three companies. *Id.* ¶ 69. It then jumps to the conclusory assertion, again "on information and belief," that "NVIDIA materially aided numerous other customers in downloading, using, and storing The Pile (and Books3) dataset." *Id.*

Plaintiffs were not diligent in raising this issue, because the video Plaintiffs rely on was publicly available more than a year before they filed their operative complaint and the scripts were likewise publicly available as early as January 2023. McCracken Decl., Ex. 10 at 3 (showing that the public webpage Plaintiffs cite on page 11 of their Motion was "[l]ast Updated on Jan 10, 2023"), Ex. 11 (showing that the script Plaintiffs cite on page 5 of their Motion was publicly available on March 18, 2024). Plaintiffs were also made aware of these scripts during discovery, when NVIDIA produced a substantively similar script on January 24, 2025, well before the April 2025 amendment deadline. *Id.* ¶ 17, Exs. 12, 13. And in any event, any "information and belief" allegations Plaintiffs make in the PFAC would also have been available to them in the original complaint.

Plaintiffs try to excuse their lack of diligence by claiming that "public information did not adequately detail how this downloading was accomplished and whether it would actually constitute a distinct infringing act." Mot. at 11. Yet the very website Plaintiffs cite—published in January 2023—states that these scripts can be used "to download and prepare the Pile dataset." McCracken Decl., Ex. 10 at 1. This website also identifies the script used to prepare the dataset, "/opt/bignlp-bignlp-scripts/main.py," which Plaintiffs could have inspected to determine the downloading process. *Id.* at 2.

Plaintiffs have also failed to plead non-conclusory facts to support claims of indirect infringement. "To state a claim for vicarious copyright infringement, a plaintiff must allege that the defendant has (1) the right and ability to supervise the infringing conduct and (2) a direct financial

1    interest in the infringing activity." *Perfect 10, Inc. v. Visa Int'l Serv. Ass'n* ("*Visa*"), 494 F.3d 788,

2    802 (9th Cir. 2007) (citation omitted). A claim for vicarious infringement also "requires an

3    underlying act of direct infringement." *Perfect 10, Inc. v. Yandex N.V.*, 962 F. Supp. 2d 1146, 1158

4    (N.D. Cal July 12, 2013). Plaintiffs' allegations about direct infringement "on information and

5    belief" are so generic and conclusory that they fail to plead plausible allegations of direct

6    infringement. *Menzel v. Scholastic, Inc.*, 2018 WL 1400386, at \*2 (N.D. Cal. Mar. 19, 2018). Nor

7    does the PFAC allege any facts that establish NVIDIA's supervision and control over its customers.

8    *Visa*, 494 F.3d at 803 (holding that a defendant's "ability to control its own [] webpages does not

9    give [that defendant] the right to control the infringing acts of third parties"). Plaintiffs' allegations

10   also fail to establish the "causal relationship between the infringing activity and any financial benefit

11   a defendant reaps" required for vicarious infringement. *Perfect 10, Inc. v. Giganews, Inc.*, 847 F.3d

12   657, 673 (9th Cir. 2017). Instead, the scripts identified by Plaintiffs were available for free, could

13   be used to create any tyspe of model on any platform, not just an NVIDIA one, and provide only a

14   substitute for downloading the same materials from Eleuther AI's website.

15          Similarly, to state a claim for contributory infringement, Plaintiffs must allege direct

16   infringement by a third party and that NVIDIA, "with knowledge of the infringing activity,

17   induce[d], cause[d] or materially contribute[d] to the infringing conduct of another." *Cobbler*

18   *Nevada, LLC v. Gonzales*, 901 F.3d 1142, 1147 (9th Cir. 2018)*.* For the reasons stated above,

19   Plaintiffs have not alleged specific acts of third-party infringement. *Concord Music Grp., Inc. v.*

20   *Anthropic PBC*, 2025 WL 1487988, at \*4 (N.D. Cal. Mar. 26, 2025) (granting motion to dismiss

21   contributory infringement claim where plaintiffs "have not clearly alleged a predicate act of direct

22   third-party infringement."). In addition, because the framework has many substantial non-infringing

23   uses (*e.g.*, training LLMs on other datasets), Plaintiffs' contributory infringement claim is futile

24   under *Sony Corp. of America v. Universal City Studios, Inc.*, 464 U.S. 417 (1984).

25          **D.  Plaintiffs' Proposed Amendments Would Disrupt the Case and Prejudice NVIDIA**

26          Plaintiffs' attempt to drastically expand the case comes after nineteen months of litigation

27   and on the eve of the close of discovery, after NVIDIA has substantially completed its document

28   production. "Expense, delay, and wear and tear on individuals and companies count toward

1   prejudice" especially when "[t]he parties have engaged in voluminous and protracted discovery."

2   *Kaplan v. Rose*, 49 F.3d 1363, 1370 (9th Cir. 1994); *Kennedy v. Nat. Balance Pet Foods, Inc.*, 2008

3   WL 2053551, at *2 (S.D. Cal. May 13, 2008) (denying leave to amend). That is the case here.

4         To date, NVIDIA has responded to over 140 written discovery requests, reviewed at least

5   300,000 documents and tens of gigabytes of data based on the operative complaint, substantially

6   completed its production of non-custodial and custodial ESI for the custodians identified prior to

7   the original substantial completion deadline (23,467 documents and 69 gigabytes of technical data),

8   litigated three motions on the appropriate scope of discovery, and developed a case strategy based

9   the operative original complaint and the Court's rulings on the scope of discovery. McCracken Decl.

10   ¶ 3. NVIDIA substantially completed production of documents for all custodians requested by

11   Plaintiffs and non-ESI documents by October 9, 2025, as set forth in the Court's original schedule.[6]

12   The proposed amendment would triple the number of claims, introduce *all* of NVIDIA's LLMS,

13   add multiple new datasets, and require extensive additional discovery and motion practice. NVIDIA,

14   and the Court, would have to redo nineteen months of litigation, starting at the beginning on new

15   claims.

16         This is exactly the type of undue prejudice that justifies denying leave to amend under Rule

17   15. *Spin Master Ltd v. Your Store Online*, 2010 WL 4883884, at *6 (C.D. Cal. Nov. 22, 2010)

18   (finding undue prejudice and denying leave to add new copyright claim because doing so "would

19   force [defendant] to engage in significant additional discovery, likely necessitating a delay in the

20   discovery cut-off" and, "[m]ore importantly, [defendant] would have to undertake the defense of a

21   whole new, tangentially related copyright claim"); *Morongo Band of Mission Indians v. Rose*, 893

22   F.2d 1074, 1079 (9th Cir. 1990) ("The new claims set forth in the amended complaint would have

23   greatly altered the nature of the litigation and would have required defendants to have undertaken,

24   at a late hour, an entirely new course of defense.").

25   _____

26   [6] The Court recently granted Plaintiffs' motion to amend the scheduling order and moved the substantial-completion deadline to November 24, 2025. Plaintiffs' inability to meet the original deadline does not alleviate the expense NVIDIA has already undertaken to abide by the Court's

27   original deadline. To the contrary, Plaintiffs' lack of diligence further contributes to the "[e]xpense, delay, and wear and tear" that "count toward prejudice." *Kaplan*, 49 F.3d at 1370.

28

Plaintiffs offer no plan for how the parties could restart the case on a vastly expanded scope and still complete fact discovery in the remaining three-and-a-half months, two of which are over the holidays. Instead, Plaintiffs' request would almost "require the Court to extend" other deadlines, including the "deadline for the completion of fact discovery . . . in order to allow time for" NVIDIA to file a motion to dismiss on Plaintiffs' new counts and otherwise address Plaintiffs' belated theories. *Watkins v. MGA Entm't, Inc.*, 574 F. Supp. 3d 747, 760 (N.D. Cal. 2021).

Plaintiffs' cited cases show the massive differences between their request and others that have been granted. For example, in *Hwa Sung Sim v. Duran*, the plaintiff, who was *pro se* at the time the original and first amended complaints were filed, sought only to substitute a named defendant for a Doe Defendant—not to add "any additional unrelated claims, allegations, or events." 2018 WL 746490, at *1–2 (E.D. Cal. Feb. 7, 2018). The plaintiff's motion also was timely under the case schedule; thus Rule 16 did not apply. *Id.* at *2.[7] Likewise, the motion for leave to amend in *PNY Techs., Inc. v. SanDisk Corp.* was timely filed—as evident from the full sentence Plaintiffs selectively cite: "[Defendant] has a heavy burden of showing prejudice given that the motion to amend the pleadings came with four months remaining for discovery *and within the time period for making a motion to amend*." 2014 WL 294855, at *5 (emphasis added). *In re Mosaic* is distinguishable because Rule 16 did not apply there, either, and there was no prejudice to defendants because discovery into newly accused LLM models would "not represent a radical shift in the nature or scope of discovery" given the significant overlap between the new LLMs and those already accused in the case. *In re Mosaic LLM Litig.*, 2025 WL 1755650, at *3 (N.D. Cal. June 25, 2025). Finally, in *Herrera v. Cnty. of San Benito*, the plaintiff discovered certain information relevant to the original claims only after the close of discovery, and the court found that "[i]f Defendant played a role in delaying the timing of Plaintiff's motion, it is difficult to argue any resulting prejudice is undue." 2025 WL 2323350, at *5 (N.D. Cal. Aug. 11, 2025).

Here, Plaintiffs' delay stems entirely from their own lack of diligence—not from any

---

[7] Plaintiffs also misquote the dicta they cite. Mot. at 13. The full sentence states that prejudice is "mitigated if discovery is still open and *there is no evidence the non-movant would be prejudiced by the timing of the proposed amendment*." 2018 WL 746490, at *2 (emphasis added).

1 conduct by NVIDIA. The information underlying Plaintiffs' proposed amendments was available

2 before Plaintiffs filed suit, often from sources they cited in their original complaint. NVIDIA did

3 not withhold information, obstruct discovery, or otherwise contribute to Plaintiffs' failure to timely

4 amend. If anything, it was Plaintiffs who failed to respond to NVIDIA's inquiries on the scope of

5 discovery made in January 2025 (McCracken Decl. ¶ 4), and it was Plaintiffs who required NVIDIA

6 and this Court to repeatedly litigate the scope of discovery. The prejudice NVIDIA now faces in

7 litigating the case anew—restarting discovery, tripling the asserted claims, and adding an unknown

8 number of LLMs and datasets, after NVIDIA had already met the substantial completion deadline

9 for the originally identified custodians—is the direct result of Plaintiffs' inexcusable delay, not any

10 action by NVIDIA.

11 **V.     CONCLUSION**

12      NVIDIA respectfully asks the Court to deny Plaintiffs' motion for leave to amend with

13 prejudice.

14

15  Dated:  November 3, 2025                Respectfully Submitted,

16

17                                 By:  /s/ Sean S. Pak
                                        QUINN EMANUEL URQUHART & SULLIVAN, LLP

18                                      Sean S. Pak (SBN 219032)

19                                      seanpak@quinnemanuel.com
                                        50 California Street, 22nd Floor
                                        San Francisco, CA 94111

20                                      Telephone: (415) 875-6600
                                        Facsimile: (415) 875-6700

21
                                        Andrew H. Schapiro (admitted *pro hac vice*)

22                                      andrewschapiro@quinnemanuel.com
                                        191 N. Wacker Drive, Suite 2700

23                                      Chicago, Illinois 60606
                                        Telephone: (312) 705-7400

24                                      Facsimile: (312) 705-4001

25                                      Alex Spiro (admitted *pro hac vice*)
                                        alexspiro@quinnemanuel.com

26                                      295 Fifth Avenue
                                        New York, NY 10016

27                                      Telephone: (212) 849-7000
                                        Facsimile: (212) 849-7100

28

1          Rachael L. McCracken (SBN 252660)
rachaelmccracken@quinnemanuel.com

2          865 South Figueroa Street, 10th Floor
Los Angeles, CA 90017

3          Telephone: (213) 443-3000
Facsimile: (213) 443-3100

4

5          *Attorneys for Defendant,*
*NVIDIA Corporation*

Master File Case No. 4:24-cv-01454-JST (SK)

NVIDIA'S OPPOSITION TO MOTION FOR LEAVE TO AMEND COMPLAINT

1    **FILER'S ATTESTATION**

2        Pursuant to Civil Local Rule 5-1(h)(3), I attest under penalty of perjury that concurrence in

3    the filing of the foregoing document has been obtained from all other signatories to this document.

4

5    Dated: November 3, 2025

6                                    By: /s/ Rachael L. McCracken
                                         Rachael L. McCracken

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28